Author's response to reviews

Title: A case-control study of physical activity patterns and risk of non-fatal myocardial infarction

Authors:
- Jian Gong (jgong@fhcrc.org)
- Hannia Campos (hcampos@hsph.harvard.edu)
- Mark Joseph A Fiecas (mfiecas@ucsd.edu)
- Stephen T McGarvey (stephen_mcgarvey@brown.edu)
- Robert Goldberg (Robert.goldberg@umassmed.edu)
- Caroline Richardson (caroli@umich.edu)
- Ana Baylin (abaylin@umich.edu)

Version: 3 Date: 22 December 2012

Author's response to reviews: see over
Dear Editor:

Thank you for considering our manuscript “A case-control study of physical activity patterns and risk of non-fatal myocardial infarction” for publication in BMC Public Health.

We thank the reviewers for their insightful comments and we have included all their suggestions in this new version of the manuscript. A point by point answer to their comments is attached.

Copyright has not been assigned for this manuscript; it has not been published in any other form and has not been submitted for publication elsewhere. None of the authors have any conflict of interest.

Sincerely yours,

Ana Baylin, MD, DrPH
Associate Professor
Reviewer's report:
Title: A case-control study of physical activity patterns and risk of non-fatal myocardial infarction
Version: 2 Date: 9 October 2012
Reviewer: Kristi L Storti

Reviewer's report:
There is some concern regarding the standardized questionnaire used to assess physical activity. While the authors indicate that the questionnaire was validated via the Harvard Step test, plasma lipids, and obesity; when the previous articles cited were pulled, this validity information was not included and many of the findings were based upon fitness not physical activity. Therefore, more information is needed on the 18 questions used to assess physical activity and the actual validity estimates.

Author’s response: We agree with the reviewer that more information needs to be provided to show the validation of the physical activity questionnaire we used. Although we did not use a “gold standard” to validate our questionnaire, in our previous articles (citation 17), we showed that people who were more active (e.g. less time in sitting and more agriculture activities) had higher fitness scores, lower LDL cholesterol levels, and lower BMI. Further, our results on physical activity and risk of non-fatal AMI are consistent with those from previous studies (i.e. physical activity is protective for AMI), providing evidence for predictive validity. Although we are aware of the difficulty of measuring physical activity with a questionnaire, the fact that the associations of physical activity with diverse outcomes in this population goes always in the expected direction is an indirect measure of validity. We added the following paragraph on page 7:
“This questionnaire was validated in a previous study of 465 people conducted in Costa Rica (ref 17). The data showed that the reported time spent on different types of daily activities using the questionnaire predicted higher fitness scores, lower LDL levels, and lower BMI’.

In the results, page 10, the authors indicate that total energy expenditure was 32.3 METs/day for cases and 33.2 METs/day for controls, but do not indicate previously how this was calculated. In Table 2, the sum is listed as energy expenditure which is contradictory. Furthermore, total energy expenditure cannot be calculated by questionnaire alone as it consists of resting metabolic rate, the thermic effect of food, and activity related energy expenditure. An estimate of "activity" related energy expenditure or total activity related energy expenditure can be calculated but not total energy expenditure. This information needs to be changed accordingly in the text.

Author’s response: We apologize for the confusion we generated by using the incorrect term. We used ‘total energy expenditure’ to refer to the daily energy spent on all physical activities listed in our standardized questionnaire. To avoid further confusion, we have changed ‘energy expenditure on physical activity’ to ‘activity-related energy expenditure’ as the reviewer suggests. We also added how we calculated all activity-related energy expenditure (i.e. sum of energy expenditure on each activity in our questionnaire) in the 1st paragraph on page 7.
With regard to the physical activity variables, were these assessed for normality as the data presented in Table 2 are means (SD). Physical activity data is usually skewed and presented as median (IQR).

**Author’s response:** We agree with the reviewer that physical activity data should be presented as median (IQR) as they are skewed in our study. We made these changes in Table 2 and the results section (1st paragraph page 10).

With regard to walking and light activity, respondents to physical activity recall questions may tend to overestimate walking and underestimate sedentary behavior.

**Author’s response:** We totally agree with the reviewer that there is a potential for recall bias on self-reported physical activity. That’s one of our study limitations and we discuss it in detail on Page 15 & 16. In any case, we expect this bias to be mostly non-differential between cases and controls, given that cases were not aware of the protective effects of physical activity. This lack of knowledge may sound weird in other populations, but in this particular population when cases were asked for the most likely cause of their heart attacks, most people answered stress (40%), followed by smoking (14%) and general mood before the MI (10%). So, since most cases do not seem to attribute their myocardial infarction to lack of physical activity the potential for differential recall bias is diminished, and non-differential recall should bias our results towards the null.
Reviewer's report
Title: A case-control study of physical activity patterns and risk of non-fatal myocardial infarction
Version: 2 Date: 4 November 2012
Reviewer: Zubair Kabir

Reviewer's report:

Gong and colleagues have examined the risk of non-fatal MI in physically inactive individuals in Costa Rica employing a population-based case-control design technique and addressing the measurement of a complex exposure level such as physical activity levels through principal component analysis (PCA) technique.

Major compulsory revisions:
Cases were first time MI survivors selected from a hospital, while controls were frequency-matched to several factors (age, sex and residence) and were selected from the population randomly using the National Census and Statistics Bureau. My main concern is the lack of detailed information on the selection procedure of controls. Just mentioning ‘randomly’ selected (1:1) using the Census is patchy and inadequate. Whether a computerized matching program or random-digit dialing technique was utilized or not to select population controls from the source population is important to assess the degree of selection bias in this study design.

Author’s response: We understand the reviewer’s concern and we are giving more details on the control selection below. The Costa Rican census was used to select the matched controls. The first step after the study cardiologist confirmed the case disease status, was to fill out a Census request. This form included the area of residence of the case “canton” (county), a target age range plus or minus 5 years of the case’s age, and gender. With this information, a computer printout with all the eligible subjects was generated. Ten subjects were randomly selected by a lottery system using the observation number in the computer printout. The specific residence of each eligible subject was located in a Census map using map segments of 60 households each. The order by which eligible homes were visited was selected by a lottery system. This lottery was carried out to avoid bias for long distance, extreme poverty, high crime area, or very wealthy neighborhoods. Furthermore, Costa Ricans are very friendly and at the time the study was carried out, were still eager to participate in scientific studies, as shown by the high participation in our study: 98% for cases and 88% for controls. Although, there is always some potential for selection bias, this has been clearly minimized in this study.

Second, recall bias and measurement bias is an inherent methodologically limitation of a case-control design. Although a standardized questionnaire and ‘trained’ interviewers were recruited for this study design to reduce recall and measurement biases, the extent of training and the numbers of interviewers trained were not reported in detail. Exposure misclassification bias is also a methodological challenge especially when self-reported information was used, despite using a previously validated questionnaire.

Author’s response: We understand the reviewer’s concern about the lack of details. We mentioned our study design and data collection process briefly because we have discussed them in detail in our previous published papers (ref. 14 & 15). To minimize
the potential for bias we did the following. All visits in our study were performed by our well-trained interviewers. The interviewers were not informed about the major hypothesis of the study. They visited cases, on average, within 3 weeks of hospital discharge (for controls, hospital discharge of the corresponding case subject) and when possible, by the same interviewer. Identical questionnaires and data collection procedures were used for cases and controls. We have added more details to the methods section (last paragraph on page 6 & 7). In addition, we discuss those limitations (potential for recall bias and measurement error) in the discussion section.

Third, income levels are not best indicators of socio-economic status and are not consistently available especially in low and middle-income nations. Educational levels capture socio-economic status better and such information was not available.

**Author’s response**: We agree with the reviewer that income levels may not be available in low and middle-income nations, but they were available in our study and they actually were better predictors of SES than education levels. In any case, we redid the analyses in Table 5 using education level instead of income, and found the results are very similar (please see the table below). Thus, we did not change that in the paper but we are happy to do so, if the reviewer considers it necessary.

<table>
<thead>
<tr>
<th>Quintiles of component scores for the first factor (rest/sleep)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>1.0</td>
<td>0.85</td>
<td>0.68, 1.06</td>
<td>0.79</td>
<td>0.64, 0.98</td>
</tr>
<tr>
<td>Model 2</td>
<td>1.0</td>
<td>0.85</td>
<td>0.68, 1.06</td>
<td>0.79</td>
<td>0.63, 0.98</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quintiles of component scores for the second factor (agricultural job)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>1.0</td>
<td>1.11</td>
<td>0.90, 1.38</td>
<td>1.20</td>
<td>0.96, 1.49</td>
</tr>
<tr>
<td>Model 2</td>
<td>1.0</td>
<td>1.13</td>
<td>0.90, 1.38</td>
<td>1.23</td>
<td>0.99, 1.53</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quintiles of component scores for the third factor (light indoor activity)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>1.0</td>
<td>0.95</td>
<td>0.76, 1.19</td>
<td>0.87</td>
<td>0.70, 1.09</td>
</tr>
<tr>
<td>Model 2</td>
<td>1.0</td>
<td>0.95</td>
<td>0.77, 1.19</td>
<td>0.87</td>
<td>0.70, 1.08</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quintiles of component scores for the fourth factor (manual labor job)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>1.0</td>
<td>0.92</td>
<td>0.74, 1.14</td>
<td>1.01</td>
<td>0.81, 1.25</td>
</tr>
<tr>
<td>Model 2</td>
<td>1.0</td>
<td>0.92</td>
<td>0.75, 1.14</td>
<td>1.01</td>
<td>0.81, 1.26</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quintiles of total activity-related energy expenditure (METs / day)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>1.0</td>
<td>0.80</td>
<td>0.64, 1.00</td>
<td>0.64</td>
<td>0.51, 0.80</td>
</tr>
<tr>
<td>Model 2</td>
<td>1.0</td>
<td>0.80</td>
<td>0.64, 1.00</td>
<td>0.64</td>
<td>0.51, 0.79</td>
</tr>
</tbody>
</table>

*a the first quintile as reference group; ^b odds ratio; ^c 95% confidence interval
Model 1: adjusted for matching factors, smoking status, **annual income**, and total saturated fat intake per day
Model 2: adjusted for matching factors, smoking status, **education levels** (Years in a formal school system: 0, 1-6, 7-14, 15+), and total saturated fat intake per day

Fourth, table 1 indicates that matching failed to achieve a comparable background characteristic between cases and controls, especially in terms of smoking and other main cardiovascular risk factors, thus questioning the selection of controls in terms of the exposure of interest.
**Author's response:** Our matching variables are age (±5 years), sex, and area of residence. The three variables are similar between cases and controls, which indicated that our matching was successful. We do observe differences between cases and controls in AMI risk factors such as smoking status, waist-to-hip ratio, hypertension, diabetes, and hypercholesterolemia (Table 1) as we should expect since we did not matched on those factors. In any case, we have adjusted for all potential confounders in our analyses. If there are some potential confounders not included in the final models is because their inclusion did not change the results and therefore were removed from the final model.

Fifth, conditional logistic regression modeling was performed simultaneously adjusting for the matching factors (age, sex and residence), which is technically inappropriate for a matched case-control study analysis, thus influencing the effect estimates.

**Author's response:** We apologize for the confusion. Actually, matching factors were not included in the models (that would be inappropriate as the reviewer suggests) but were taking into account in the analysis by conditioning on them by using the STRATA statement in PROC LOGISTIC in SAS. The footnote for model 1 in Table 5 is used to show readers that the results of model 1 are based on the matched case and control pairs on age, sex and residence.

Sixth, weighting of MET variables for each component of physical inactivity level measured needs further elaboration.

**Author's response:** We agree with the reviewer that we need to explain more details on the weightings/loadings in each component. We added the direction and more detailed information of the loadings for each component in the last paragraph on page 10.

Seventh, a U-shaped /J-shaped exposure-response association employing a case-control study design should be interpreted with caution, especially when a temporal sequence is unclear.

**Author's response:** We appreciate the reviewer's reminder that there is one more limitation in our study. In a case-controlled design, the temporal relationship between exposure and outcome is unclear. Accordingly, we cannot draw a conclusion on the causal relationship between extracted physical activity patterns and AMI risk. We added this limitation in the discussion section on page 15.

Finally, PCA for deriving physical activity levels is relatively novel and needs to be reproduced in different population settings for validity and precision, especially when individuals are studied.

**Author's response:** We totally agree with the reviewer that PCA on physical activity research needs to be validated in other populations. So we added the following sentence in the last paragraph on page 16: ‘Further research on different populations is required to validate the application of PCA to deriving physical activity patterns and confirm our findings.’
Reviewer's report
Title: A case-control study of physical activity patterns and risk of non-fatal myocardial infarction
Version: 2 Date: 9 November 2012
Reviewer: Wojciech Kazimierz
Reviewer's report:

This is an interesting study analyzing the association between various physical activity patterns and the risk of acute myocardial infarction in a large sample of inhabitants of Costa Rica.

The method of principal component analysis related to physical activity assessment could be useful for other authors and similar studies. Results of Jian Gong et al study showing U-shaped association between rest/sleep pattern and risk of AMI is of great practical importance similarly as data demonstrating plateau between total energy expenditure and AMI risk at high levels of physical activity. There are relatively few studies analyzing association between various physical activity domains and cardiac risk especially from countries outside Europe and North America.

In my opinion several minor essential revisions or additional comments are needed before printing the paper in BMC Public Health:

1. The authors use a statistical method "natural cubic spline models" not widely used in similar studies analyzing relationship between risk factors and AMI incidence. Thus I think more clear explanation is needed to justify the use of this statistical method. Probably the method description presented in the paper (page 8-9) is not clear for most readers being not professional statisticians.

Author’s response: We appreciate the reviewer’s suggestion. We added the following sentence on page 9 to further explain natural cubic splines: “Natural cubic splines are smooth polynomial functions that can be used to fit data and accommodate potential changes in the direction of the association across the distribution of an exposure. They are useful to examine non-parametrically the potential non-linear relation between the exposure and the outcome of interest.”

2. For many readers of the BMC Public Health especially from Europe and Asia, the health situation in Costa Rica is not known, thus a short description of the mortality reasons due to CVD diseases and recent trends of AMI mortality would be useful.

Author’s response: According to the reviewer’s suggestion, we added the following sentence on page 6 to briefly describe the mortality due to CVD in Costa Rica: ‘In Costa Rica, CVD has been the country’s leading cause of death since 1970 and the mortality rate for CVD has been declining since 2002 according to 2007 Health in the Americas, a report from World Health Organization.”.

3. The numbers given in Table 3 are not enough clear for me and should be better described. Probably a short description of one row components – for example for work
in construction - describing precisely what the positive and negative numbers really mean would be helpful.

**Author’s response:** We apologize for not clearly explaining the results in Table 3. The numbers in each column are the weightings/loadings of each type of physical activity on the corresponding component or physical activity pattern and ‘positive’ and ‘negative’ represent the direction of the association between physical activity and components. We added the direction and more detailed information of the weightings/loadings for each component in the last paragraph on page 10. We have also added the following example as a footnote to the table to clarify the interpretation of the loadings: “The patterns are named based on activities that have high positive loadings, for example the agricultural job pattern have high positive loadings on activities like “standing and squatting in the garden work” or “work in agriculture”, while it has high negative loadings in activities that are more representative of other patterns (i.e. “Stand in very light activities at work or at home such as filing, coping, and doing laundry”)

The overall assessment of the whole study is positive and I think the paper should be published after minor corrections/complementary information.

**Author’s response:** we really appreciate the reviewer’s comments.